Estimation of soil moisture using modified Antecedent Precipitation Index with application in landslide predictions

Binru Zhao[1,2], Qiang Dai[3], Dawei Han[2], Huichao Dai[1], Jingqiao Mao[1], Lu Zhuo[2], Guiwen Rong[4]

[1]College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, China.

[2]Department of Civil Engineering, University of Bristol, Bristol, UK.

[3]Key Laboratory of VGE of Ministry of Education, Nanjing Normal University, Nanjing, China.

[4]College of Earth and Environment, Anhui University of Science and Technology, Huainan, China

Corresponding author: Qiang Dai (bz17336@bristol.ac.uk)

Abstract

Soil moisture plays a key role in land-atmosphere interaction systems. Although it can be estimated through in-situ measurements, satellite remote sensing and hydrological modelling, using indicators to index soil moisture conditions is another useful way. In this study, one of these indicators, The Antecedent Precipitation Index (API) is explored. Modifications were proposed to the conventional version of API by introducing two parameters to make it more in line with the physical process. First, the recession coefficient is allowed to vary with the change of air temperature, which could take into account the variation of the evapotranspiration process. Second, the API value is restricted by the maximum value of API, accounting for the maximum water holding capacity of the soil. The modified API was then calibrated and validated by comparing with the in-situ measured soil moisture. The better correlation between these two datasets demonstrates that the modified API could better indicate soil moisture conditions, compared with the conventional API. The capability of the modified API to index soil moisture conditions was further explored by applying it to landslide predictions in the Emilia-Romagna region, northern Italy. Here the recent 3-day rainfall vs the antecedent soil wetness thresholds (RS thresholds) were constructed, in which the soil wetness is indexed by the modified API. The validation of RS thresholds was carried out with the use of the contingency matrix and Receiver Operating Characteristic (ROC) curves. By comparing the prediction performance between RS thresholds and rainfall thresholds, it is found that RS threshold could provide better prediction capabilities in terms of higher hit rate and lower false alarm rate. The positive results indicate that the modified API could provide superior performance of indexing soil moisture conditions, demonstrating the effectiveness of the proposed modifications.

1 Introduction

Soil moisture in this study refers to the water content in the unsaturated zone. It plays a crucial role in the land-atmosphere interaction, through governing the water and energy balance between the land surface and the first layer of the atmosphere. As a result, the estimation of soil moisture is important in many scientific and practical issues. For rainfall-runoff modeling, the soil moisture condition prior to rainfall storms has been recognized as a key factor in determining the catchment runoff response (Brocca et al. 2010, Castillo et al. 2003, Koster et al. 2010). For numerical weather prediction and climate modelling, soil moisture is a major consideration due to its role in governing the partitioning of the mass and energy fluxes between the land and atmosphere (Bolten et al. 2010, Koster et al. 2004, Koster et al. 2009). The antecedent soil moisture is also known as an important factor in the initiation of rainfall-triggered landslides (Glade et al. 2000, Zêzere et al. 2015).

Soil moisture estimates can be obtained in different ways, such as in-situ measurements, remote sensing and hydrological modelling. In-situ measurements are arguably the most accurate estimation of soil moisture; however, point-based measurements make it limited in terms of spatial extent (Brocca et al.

2007). Due to the high cost of installation and maintenance, the in-situ measurements are not always available in the interested areas. In previous studies, the most common use of the in-situ measured soil moisture is to calibrate and test other estimates of soil moisture. Remote sensing technology has been widely used to estimate surface soil moisture in recent years (Entekhabi et al. 2010, Kerr et al. 2010), including the Soil Moisture and Ocean Salinity (SMOS) satellite launched by European Space Agency (SEA) and the Soil Moisture Active and Passive (SMAP) program scheduled by National Aeronautics and Space Administration (NASA). Many studies have evaluated and validated the remote sensed soil moisture products by comparing them with in-situ measurements (Draper et al. 2009, Gruhier et al. 2009, Jackson et al. 2010, Wagner et al. 2006). They found the remote sensed soil moisture could capture soil moisture temporal variations in good agreement with in-situ measurements. The remote sensing products can provide quantitative soil moisture information at a global scale with free availabilities, and have been applied to many hydrological, meteorological and agriculture applications, despite the coarse resolution. Some attempts have also been made to estimate soil moisture with the use of hydrological modelling. Posner and Georgakakos (2015) utilized spatially distributed operational hydrological models to estimate depth-integrated soil moisture, and then applied it to a regional forecasting system for landslide hazard threat level in EI Salvador. Valenzuela et al. (2017) analyzed soil moisture conditions of 84 landslides in Asturias, NW Spain. The soil moisture was represented with the index Available Water Capacity (AWC), which is extracted from daily water balance models. Although the model-based method is a useful way to estimate soil moisture conditions, it has a high demand for data inputs and normally computationally intensive especially for larger study areas.

In addition to the aforementioned conventional methods, some indicators are also used as a means of estimating soil moisture. Antecedent Precipitation Index (API) is one of these indicators, proposed by Linsley et al. (1949). API is based on precipitation that has occurred over the preceding days, and due to the easier availability of the precipitation observations, the use of API is more practical for some applications where general indications of soil moisture conditions can meet demand. Crozier and Eyles (1980) employed this index to characterize the effect of antecedent soil moisture conditions on the occurrence of rainfall-triggered landslides. Crow et al. (2005) explored the effect of antecedent soil wetness on runoff forecasting with the use of API. Although API is considered as a useful indicator of soil moisture and easier to use in practical applications, deep investigations on its usage are still absent, and there are some questions remaining unexplored. For example, API is derived from the antecedent precipitation with a recession coefficient representing the rates of drainage and evapotranspiration processes. The choice of the length of the preceding period and the recession coefficient is unclear. The preceding period chosen as significant differs considerably in the previous studies (Crozier and Eyles 1980, Zêzere et al. 2005), varying from 10 days to 60 days. As for the determination of the recession coefficient, most studies used the value of 0.84, which comes from Ottawa (United States) streamflow data in the work of Crozier and Eyles (1980). When defining the landslide-triggering rainfall thresholds using the index for antecedent rainfall, Glade, et al. (2000) derived the recession coefficient from the recession curves of storm hydrographs for each region. The calculated thresholds show regional differences in susceptibility of a given landscape to landslide-triggering rainfall. Furthermore, there are limited investigations directly exploring the performance of API in indicating soil moisture conditions, due to the lack of in-situ measurements as the benchmark. Filling these knowledge gaps will benefit the applications of API to a wider range of practical problems.

Based on the above, the aim of this study is to explore the improved usage of API to index soil moisture conditions. It is also attempted to modify API to make it more in line with the physical process. From a hydrological point of view, there are two aspects worth improving for API. First, in the definition of API, the recession coefficient is assumed to be constant throughout the year, ignoring the variation of the evapotranspiration process, which may be related to some factors such as temperature, wind speed, relative humidity, etc. Second, the API expression lacks the consideration of the maximum water capacity of the soil layer, which may cause an overestimation of the soil moisture. Therefore, this study intends to improve the formulation of API by incorporating more considerations of the hydrological process. The performance of API in indicating soil moisture conditions is evaluated by comparing with the in-situ measurements of volumetric water content. The application of the modified API in landslide predictions is also investigated, where the antecedent soil moisture conditions before the landslide occurrences are indexed by the modified API. This study was carried out in a northern Italian region called Emilia Romagna, owing to the availability of ample landslide records and the hydrometeorological data.

2 Study Area and Data Sources

2.1 Study area

The Emilia-Romagna region is located in the north of Italy, bordered by Apennines mountains (on the south and the west), Adriatic Sea (on the east) and Po River (on the north). There is a wide flat area in the northern and eastern portions of the region, while its southern and western areas are characterized by hills and mountains, whose maximum altitude is 2165m (Figure 1). This region has a typical Mediterranean climate: summer, from approximately May to October, is warm and dry, while winter from November to April is mild/cold and wet.

The mountainous part of the Emilia-Romagna region is extremely prone to landslides. There are a variety of landslide topologies (Martelloni et al. 2011), like the rotational-translational slides, slow earth flows, complex movements, rapid shallow landslides, etc. Although the occurrence of landslides is a result of multiple factors, in the Emilia-Romagna region, the main triggering factor of landslides is rainfall. Short but intense rainfalls are more likely to trigger debris flows and shallow landslides, while deep-seated landslides and earthflows are mainly caused by moderate but prolonged periods of rainfalls (Ibsen and Casagli 2004).

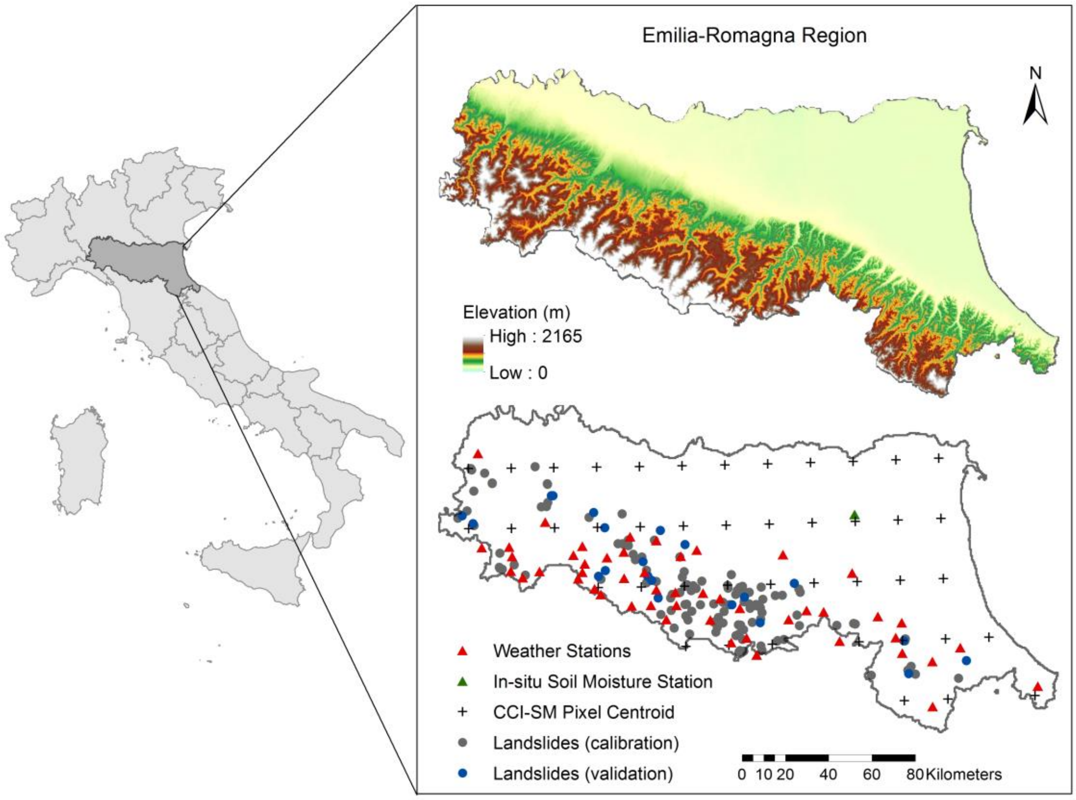

Figure 1. Location of the Emilia-Romagna region as well as the location of landslides, in-situ measurement stations and the CCI-SM centroid pixels.

2.2 Data sources

In the study area there is a hydro-meteorological network maintained by Regional Agency for the Prevention, Environment and Energy of Emilia-Romagna (Arpae), which is able to provide a variety of observations at different temporal scales, such as rainfall, pressure, air temperature, relative humidity, wind speed, soil moisture, etc. All these data can be obtained online (http://www.smr.arpa.emr.it/dext3r/). The data used in this study was extracted the here, including the in-situ measured soil moisture, rainfall, and temperature data.

There are 19 in-situ soil moisture measurement sites within the study area, where Time Domain Reflectometry (TDR) equipped with dataloggers is used to measure soil water content at different soil depths. Among these sites, only the San Pietro Capofiume site (marked with the green triangle in Figure 1) could provide long-term surface soil moisture observations (at 10 cm soil depth). Therefore, this site was selected to explore the performance of API, where data of the daily average temperature, daily cumulative rainfall and soil water content are needed. Considering the completeness of the required data, data of the period from 2006 to 2016 were extracted and used for the analysis.

For the purpose of applying the modified API to landslide predictions, the modified API was calculated for the landslide-prone area of the Emilia-Romagna region, and its validation needs to be evaluated using the in-situ soil moisture measurements. However, due to the lack of long-term in-situ measured soil moisture in this area, remote sensed soil moisture was utilized as a proxy. The modified API was calculated based on the data from 50 weather stations (marked with red triangles in Figure 1) for the period from 2006 to 2016. The remoted sensed soil moisture adopted in this study is the state-of-the-art ESA Climate Change Initiative (CCI) soil moisture product (CCI-SM hereafter). This product is produced by merging information from multiple active and passive microwave sensors, including three harmonized satellite soil moisture datasets: a merged ACTIVE (1991-2016), a merged PASSIVE (1978-2016) and a COMBINED (1978-2016). The soil moisture information provided by the CCI-SM product is the volumetric water content (m^3/m^3), with a daily temporal resolution, and 0.25-degree spatial resolution. In this study, the latest version (v04.2, released in early 2018) of CCI-SM COMBINED product was employed, whose pixel centroids are shown in Figure 1.

The landslide data was collected from Emilia-Romagna Geological Survey, an agency maintaining a catalogue of historical landslides in the Emilia-Romagna region. The landslides recorded in this catalogue were from various sources, such as reports to local authorities, national and local press, technical documents. Most landslides that led to casualties and damage were recorded, while those with little influence or damage were more likely to be undetected. In general, a range of landslide occurrence information should be gathered, such as location, date, accuracy level of the record, characteristics (length, width, type and material), triggering factors, damage and references. However, in practice, it is difficult to collect and record all the above information. For most landslides, only the occurrence location and date were recorded. Despite such a fact, this catalogue is the most complete and detailed records of landslides in the Emilia-Romagna region, and regarded as a proxy of actual landslides (Rossi et al. 2010). In this study, only the landslides with daily accuracy in terms of the occurrence date were selected for the landslide prediction analysis, with a total of 140 (Figure 1). The landslides occurred during the period from 2006 to 2014 were used to establish the thresholds for landslide occurrence, and the landslides in the period from 2015 to 2016 were for the validation of the thresholds.

3 Methods

3.1 Antecedent Precipitation Index

Antecedent Precipitation Index (API) is an index derived from the preceding daily rainfall, regarded as a simple surrogate measure of soil moisture. One common definition of API was proposed by Fedora (1987) to simulate storm hydrographs in the Oregon Coast Range, written as:

$$\mathrm{API_t} = \mathrm{k\,API_{t-\Delta t}} + \mathrm{P_{\Delta t}} \tag{1}$$

where $\mathrm{API_t}$ is the API at time t, $\mathrm{P_{\Delta t}}$ is the cumulative precipitation during the period from $\mathrm{t} - \Delta\mathrm{t}$ to t (in this study $\Delta\mathrm{t} = 1$ day), and k is the recession coefficient, which is assumed constant throughout the year.

From a hydrological point of view, there are two aspects worth improving for the above formulation of API. First, assuming the recession coefficient is a constant is not in agreement with the physical process. The recession coefficient is adopted to characterize the rate of water loss, which is a result of the drainage and evapotranspiration processes. Considering that the evapotranspiration process is dependent on multiple factors (e.g. air temperature, wind speed, relative humidity) and these factors vary through the year, assuming the recession coefficient is a constant ignores the variation of these factors and their effects on soil moisture conditions. The effect of temperature on the soil moisture evolution can be found in Figure 2, which illustrates the temporal evolution of in-situ soil moisture measurements in the San Pietro Capofiume site as well as the corresponding rainfall and temperature series. As can be seen, for three dry periods (December of 2015, March and April of 2016 and July of 2016), their rainfall conditions are similar, with little or zero amount. However, their soil moisture conditions show significant differences. For December of 2015, the soil remains in a wetter condition, while for other two dry periods, the water content is lower. This can be explained by the difference of the temperature during these periods. Higher temperature conditions benefit the evapotranspiration process and lead to a more loss of water, thus less water is attributed to the soil resulting in a lower water content. On the contrary, lower temperature conditions will result in the wetter soil conditions due to the reduced water loss. Therefore, it is necessary to take into account the variation of some factors that affect the evapotranspiration process. Due to the easier availability of temperature data, only temperature is considered in this study. For this purpose, the formulation of API expressed in Equation (1) is modified by allowing the recession coefficient to vary according to the change of temperature. As a line relationship is simpler and easy to implement, the variation of the recession coefficient is assumed as linear in this study:

$$\mathrm{k} = 0.84 + \delta\,(20 - \mathrm{T_{ave}}) \tag{2}$$

where $\mathrm{T_{ave}}$ is the daily average temperature (℃) and δ is a sensitivity parameter ($℃^{-1}$). When δ is equal to 0, the recession coefficient is constant as 0.84, which is widely used in the previous studies, recommended by Crozier and Eyles (1980). The reason for using 20 ℃ as the basis is that it is the most common temperature when the value of 0.84 is used.

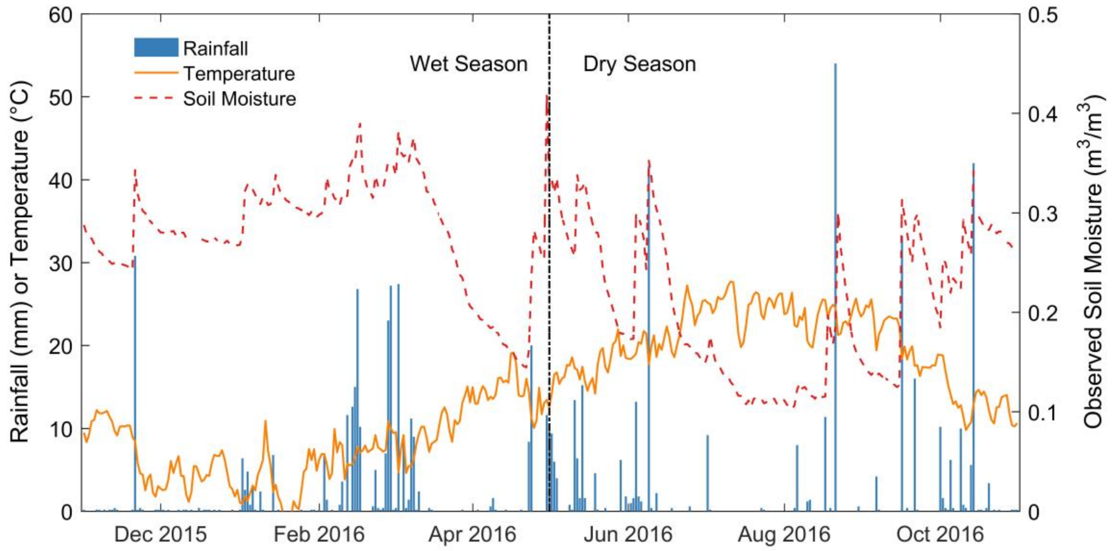

Figure 2. Time series of in-situ soil moisture measurements in the San Pietro Capofiume site as well as the corresponding rainfall and temperature series.

Second, the API expression lacks the consideration of the maximum water capacity of the soil layer. In the hydrological process, prolonged rainfall will saturate the soil allowing no additional water to be held. As a result, any additional rainfall that falls becomes overland flow. This process is termed as saturation excess overland flow. In this case, API calculated with Equation (1) will overestimate the water content in the soil. Therefore, a parameter $\mathrm{API}_{\mathrm{max}}$ is introduced to take into account the process of saturation excess overland flow: when the value of API exceeds $\mathrm{API}_{\mathrm{max}}$, it is equal to $\mathrm{API}_{\mathrm{max}}$.

The optimization of the parameter δ and $\mathrm{API}_{\mathrm{max}}$ is carried out by comparing the modified API with the observed soil moisture for the period from 2006 to 2013. As the modified API attempts to capture soil moisture temporal variations in good agreement with in-situ measurements, Pearson correlation coefficient is used as the evaluation criterion to measure the linear correlation between the modified API and the measured soil moisture. Pearson correlation coefficient ranges from -1 to 1, where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. The optimized parameters are then validated for the period from 2014 to 2016. Due to the lack of in-situ measured soil moisture, it is difficult to determine the optimal parameters for other locations. It is assumed that it is feasible to extrapolate the parameters at the San Pietro Capofiume site to the whole study area. In order to assess the reliability of this assumption, the modified API series is compared with the CCI-SM product for each weather station.

3.2 Rainfall versus soil wetness threshold

A recent rainfall versus antecedent soil wetness threshold (hereafter RS threshold) is employed to explore the application of API in landslide prediction studies. The soil wetness here is the quantification of the soil moisture condition. RS threshold consists of two components: one is the recent 3-day cumulative rainfall (R) before the landslide occurrence, and the other is the antecedent soil wetness of the day preceding the recent 3-day (S), which is indexed by the modified API as proposed. The modified API here is scaled with the maximum and minimum value, thus ranging from 0 to 1 with higher values corresponding to wetter soil conditions.

For the purpose of constructing and testing RS threshold, all datasets are divided into two periods: 2006-2014 for the construction of RS threshold, and 2015-2016 for the evaluation of the threshold. RS threshold is determined by various combinations of the critical value of landslides' rainfall and soil wetness, which are defined with their different percentiles. Taking the rainfall's 5th percentile (P5) as an example, it means there are 5% landslides with the recent 3-day cumulative rainfall value less than P5. The two components of the RS threshold are used separately in this study: the critical value of the antecedent soil wetness is firstly used as a criterion and then the critical value of rainfall is used. The reason of not constructing the functional relation between these two components (like the power law relation between rainfall intensity and rainfall duration in the rainfall threshold) is because this relationship remains unknown, although there are some studies assuming it as the linear relation (Chleborad et al. 2008, Mirus et al. 2018, Scheevel et al. 2017). The landslide occurrence is predicted only when these two components' critical values are exceeded. The prediction performance of different thresholds is evaluated with the help of the contingency matrix and Receiver Operating Characteristic (ROC) curves. This is the most common manner used in landslide early warning studies (Gariano et al. 2015, Mirus, et al. 2018, Staley et al. 2013).

Hit Rate (HR) is also known as the true positive rate, and used to measure the proportion of landslides that are correctly predicted:

$$\mathrm{HR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

False Alarm Rate (FAR) is also known as the false positive rate, and used to measure the proportion of false alarms over the events when no landslide occurs:

$$\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{4}$$

In Equations (3) and (4), True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) are four possible outcomes of the thresholds' prediction results. TP means the threshold predicts landslide occurrences successfully; FN is an error where the threshold does not predict the occurrence of landslides; however, in reality landslides occur; FP is an error where the threshold predicts the occurs of landslides; however, there is no landslide occurrence in reality; TN means the threshold correctly predicts the non-occurrence of landslides.

The value of HR and FAR ranges between 0 and 1. When HR is equal to 1 and FAR is equal to 0, the optimal performance is achieved. This is referred to a perfect point. For the better measurement of the gap to the perfect point, the Euclidean distance (d) is also calculated for each threshold scenario. The smaller the distance, the better the prediction performance.

$$\mathrm{d} = \sqrt{(\mathrm{FAR})^2 + (\mathrm{HR} - 1)^2} \tag{5}$$

3.3 Rainfall threshold

In order to directly compare the prediction performance of RS threshold with that of the rainfall threshold, the cumulative event rainfall E (mm) versus rainfall duration D (day) threshold was also constructed using the Frequentist approach proposed by Brunetti et al. (2010). They assumed the general formulation of threshold curves as a power law:

$$\mathrm{E} = \alpha \cdot \mathrm{D}^{\gamma} \tag{5}$$

where α is a scaling constant (the intercept at the value of D equal to 1), γ is the shape parameter (defining the slope of the power law curve). The cumulative rainfall E (mm) and rainfall duration D (day) are calculated based on rainfall events, which are identified using the automatic procedure proposed by Melillo et al. (2014). Thresholds with different percentiles are calculated and evaluated. The data used for the construction and test of rainfall thresholds are the same as that for RS thresholds, and the evaluation method also remains the same.

4 Results

4.1 The effect of the initial value

Before using Equation (1) to calculate API, it was necessary to explore the effect of the initial value and the determination of the period length for recursion, as the expression of API in Equation (1) is a recursive form. For this purpose, different initial values were designed ranging from 0 to 100 mm. Once the initial value is given, the expression in Equation (1) was run for the next 200 days. The temporal evolution of API with different initial values is shown in Figure 3a. It is clear to see whatever the initial value is, the value of API after near 60 days remains the same, although there is a distinct difference in the first 20 days. In other words, the effect of initial value decreases and becomes insignificant after the 60th day. In order to exclude the influence of the position of the initial day, the above procedure was repeated by choosing different dates as the initial day. Here 366 days of the year 2012 were used for analysis. It is found that the day no longer affected by the initial value distributes in the range from the 57th to 65th day, with the median value as the 59th day (Figure 3b). Based on the above results, API was calculated with the initial period of 60 days, after which the initial value has no longer effect on the API value. In this study, 30 mm was used as the initial value, which is the average level of API.

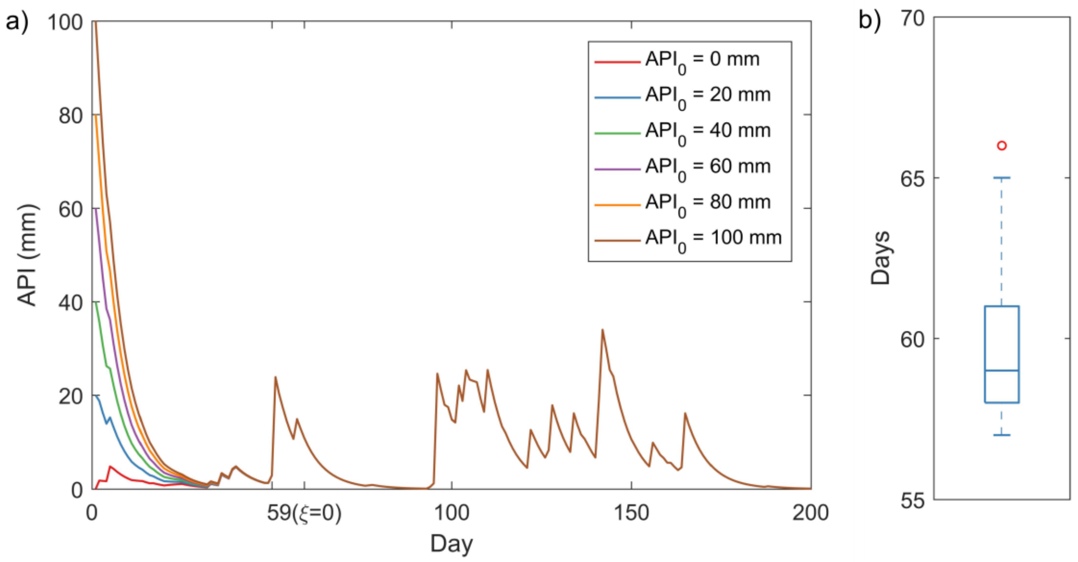

Figure 3. The effect of the initial value, a) time series of API with different initial values, b) the distribution of the day no longer affected by the initial value.

4.2 The modified API

Two parameters δ and API_{max} were introduced to modify API. To optimize these two parameters, different combinations of δ and API_{max} were tested. Their performance was evaluated with the use of Pearson correlation coefficient (r), which w calculated using the modified API series and the observed soil moisture series of the period from 2006 to 2013. To make the recession coefficient greater than 0.5, considering the variation range of the temperature, the sensitive parameter δ ranging from 0 to 0.03 was explored. As for API_{max}, if a too small value is used as API_{max}, API remains the same as API_{max} for most cases, which is not consistent with the reality. Here 10 mm was considered as the lower limit of the variation range. Given the maximum value of API, the upper limit of the variation range was determined as 200 mm. Some representative results are showed in Figure 4. As API_{max} is regarded as an indirect measure of water capacity of the soil, it should remain constant for a particular type of soil. Therefore, the value of API_{max} is firstly determined for the study site. It is interesting to find that for all values of δ, the correlation coefficient between soil moisture and API has the best value when API_{max} is around 35 mm. Therefore, 35 mm is selected as the optimal value of API_{max}. In the case of API_{max} as 35 mm, it is clear that the correlation coefficient increases greatly when δ changes from 0 to greater than 0, indicating that the performance of API could be improved by allowing the recession coefficient to vary with the temperature. The improvement is obvious when δ changes from 0 to 0.01, with the correlation coefficient increasing from around 0.55 to near 0.9. However, after δ larger than 0.01, increasing the value of δ no longer results in the significant improvements and even worsens of the correlation coefficient, and the optimal result is reached at δ as 0.012. As a result, the optimal value of δ is selected as 0.012.

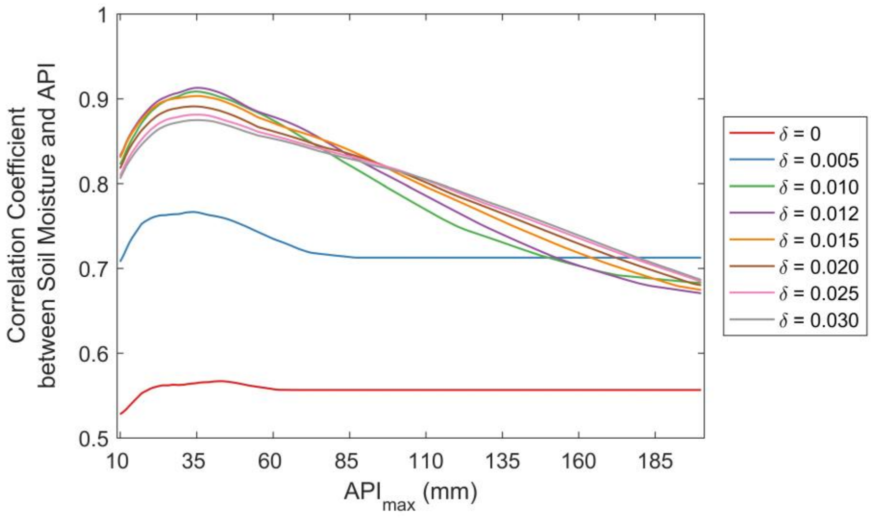

Figure 4. The optimization of parameters δ and $\mathrm{API_{max}}$.

To validate the optimized parameters, an evaluation of the modified API during the independent period from 2014 to 2016 was carried out. The scatter plot of API against the observed soil moisture as well as the fitted curve is shown in Figure 5 for both versions of API. Hereafter, the conventional API is represented by $\mathrm{API_c}$, and the modified API is represented by $\mathrm{API_m}$. From Figure 5a, the linear correlation relationship between the observed soil moisture and $\mathrm{API_c}$ is insignificant, and it is found that a power function has the best fit to the data points. As for Figure 5b, $\mathrm{API_m}$ has a significant linear positive relationship with observed soil moisture values, with Pearson correlation coefficient increasing from 0.51 to 0.88. This not only indicates that the calibrated parameters' performance is reliable for the independent period, but also demonstrates that introducing two parameters to $\mathrm{API_c}$ really improves the API's performance of indicating soil moisture conditions.

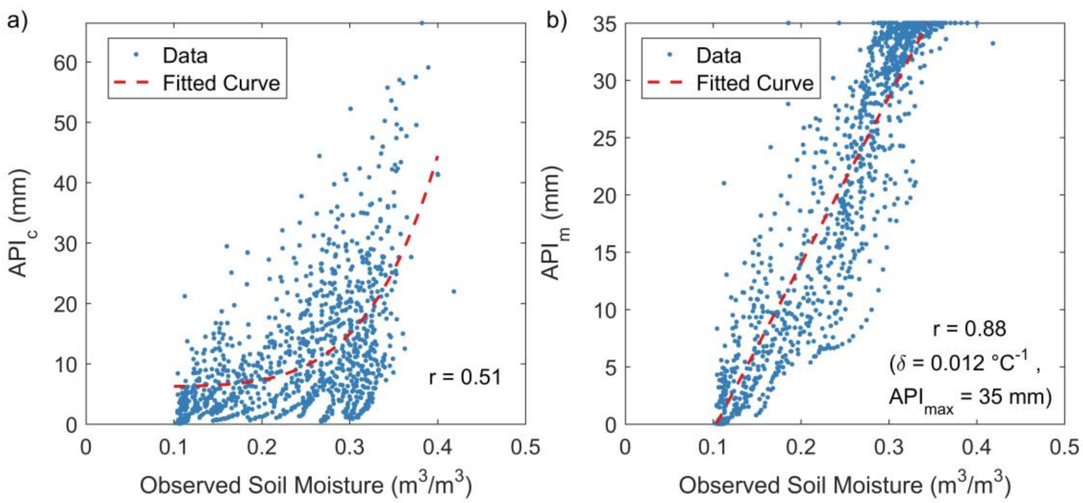

Figure 5. The scatter plot of API against the observed soil moisture as well as the fitted curve, a) for $\mathrm{API_c}$, b) for $\mathrm{API_m}$.

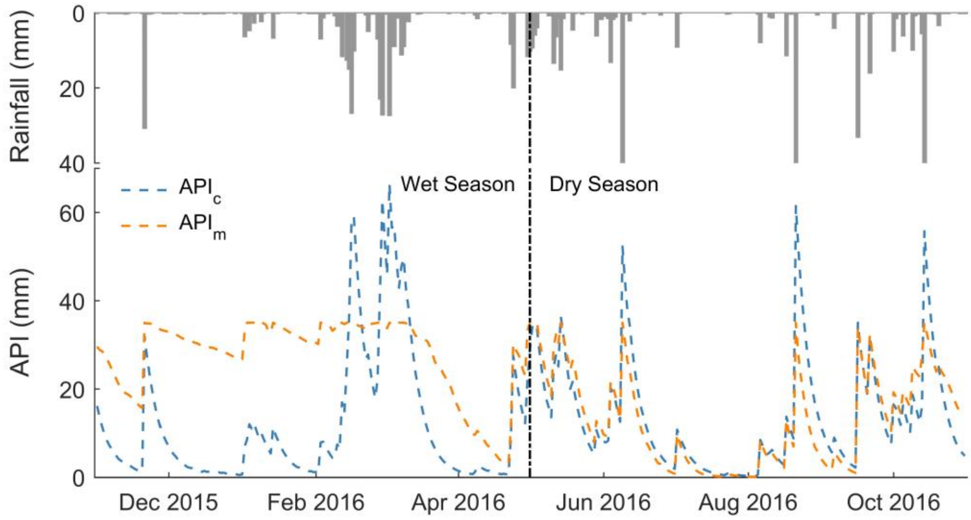

Figure 6. Time series of API_c and API_m as well as rainfall for the period from November 2015 to October 2016.

For better demonstrating the superior performance of API_m, the temporal evolution of API_c and API_m as well as rainfall are illustrated in Figure 6. As expected, the change pattern in terms of increase or decrease is the same for these two variables, because the decrease or increase of API is mainly related to the variation of rainfall. For instance, during the period with rainfall, the soil conditions become wetter, represented by an increase in the value of API. As for the period with little or no rainfall, the soil conditions become drier, represented by a decrease in their values. Despite the same change direction, API_c and API_m show distinct differences in terms of change degree and change range. The change degree is more dependent on the contribution of the antecedent rainfall, while the change range is related to the maximum water capacity of the soil. For the wet season from November 2015 to April 2016, the difference between API_c and API_m is distinct, which could be explained by the two parameters introduced to API_m. The parameter δ leads to a higher k owing to the lower temperature in the wet season, resulting in a more contribution of the antecedent rainfall. Therefore, for the period with little or no rainfall, the value of API_m will not decrease as quickly as the API_c. As for the period with intense rainfall, the value of API_m is restricted by API_{max} , leading to the difference in terms of change range. For the dry season from May 2016 to October 2016, the biggest difference between API_c and API_m is the change range caused by the introduction of API_{max}. As for the change degree that is influenced by δ, there is no distinct difference. The reason is that during this period, the temperature is near or a little greater than 20°C, therefore the k value is similar in both expressions of API_m and API_c, and thus they have similar change degree.

4.3 Parameter Extrapolation

To apply the modified version of API to landslide prediction studies, the values of API_m are required, whose calculation requires the determination of the two added parameters. However, due to the lack of in-situ measured soil moisture in the landslide-prone area, it is difficult to optimize them in the same way as that used in the San Pietro Capofiume site. Therefore, it is assumed that it is feasible to extrapolate the parameters at the San Pietro Capofiume site to the whole study area. In order to validate the parameter extrapolation, the $\mathrm{API_m}$ series was assessed with the CCI-SM product for each weather station. Here Pearson correlation coefficient was used as the criterion. The reason for not using CCI-SM to calibrate parameters for the study area is that there are uncertainties associated with the satellite data, which will lead to more uncertainties to the determination of parameters and the calculation of $\mathrm{API_m}$. Therefore, the CCI-SM product is only used for evaluating the parameter extrapolation.

Before using CCI-SM product, its reliability and accuracy were firstly investigated. The CCI-SM series is compared with the in-situ measured soil moisture in the San Pietro Capofiume site for the period from 2006 to 2016, which are shown in Figure 7. It can be seen from Figure 7a, CCI-SM is able to capture the overall seasonal and temporal variations of soil moisture. Despite this, it is noted that there are periods that CCI-SM product shows wetter conditions than the observed. Figure 7b presents the scatter plot of CCI-SM against the observed soil moisture. The variation range for both datasets are similar, CCI-SM ranges between 0.1 and 0.4 m^3/m^3, and the observed soil moisture varies from 0.05 to 0.43 m^3/m^3. Moreover, although for some dry conditions (e.g., the observed soil moisture is between 0.05 to 0.25), CCI-SM product overestimates the soil moisture, in general, the data are distributed mainly around the identical line. The Pearson correlation coefficient of 0.68 also indicates the CCI-SM is generally in line with the in-situ measurements. In summary, despite some drawbacks of CCI-SM, it is considered acceptable to represent the temporal evolution of soil moisture conditions and can be used for the validation of the parameter extrapolation.

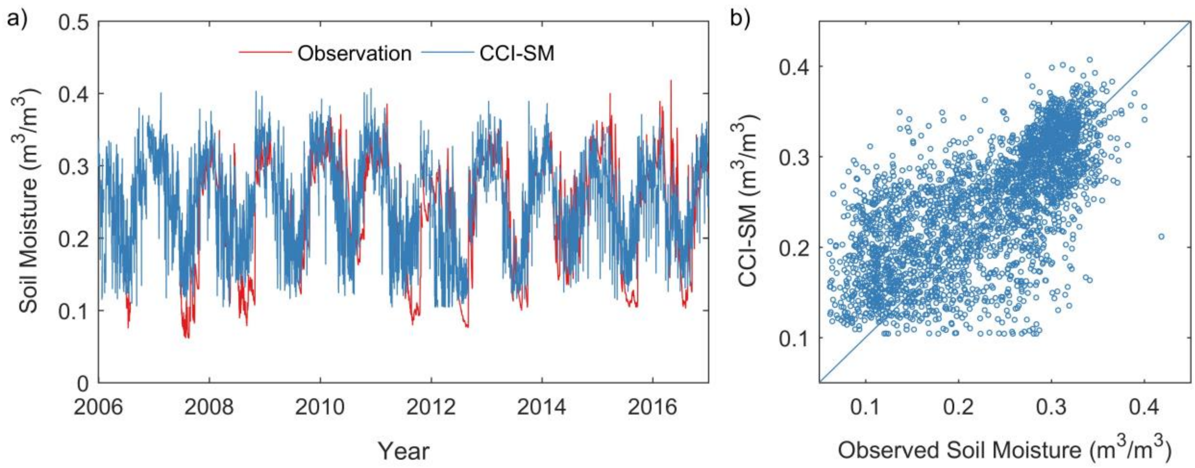

Figure 7. Comparison between the observed soil moisture and CCI-SM product, a) the time series plot, b) the scatter plot with the identical line.

The validation of the parameter extrapolation was carried out for 50 weather stations within the study area. For each station, API_m of the period from 2006 to 2016 was calculated based on its rainfall and temperature datasets, with the use of those two parameters optimized in the San Pietro Capofiume site. By matching the nearest CCI-SM pixel to the station, the CCI-SM dataset can be extracted. With the API_m and CCI-SM datasets, the Pearson correlation coefficient was calculated as the evaluation criterion. For the purpose of comparison, the same process was also carried out for API_c. The value of correlation coefficient is shown in Figure 8. As is seen, the performance of the parameter extrapolation varies with stations, with the minimum as 0.53 and the maximum as 0.78. In spite of the variance of the performance, API_m shows a great improvement over API_c by comparing the mean value of their correlation coefficients (growing from 0.48 to 0.70). Although the value of the correlation coefficient is not great enough, given the uncertainties associated with the CCI-SM product, the performance of the parameter extrapolation is regarded as acceptable and API_m can be used as an indicator for soil moisture conditions in the study area.

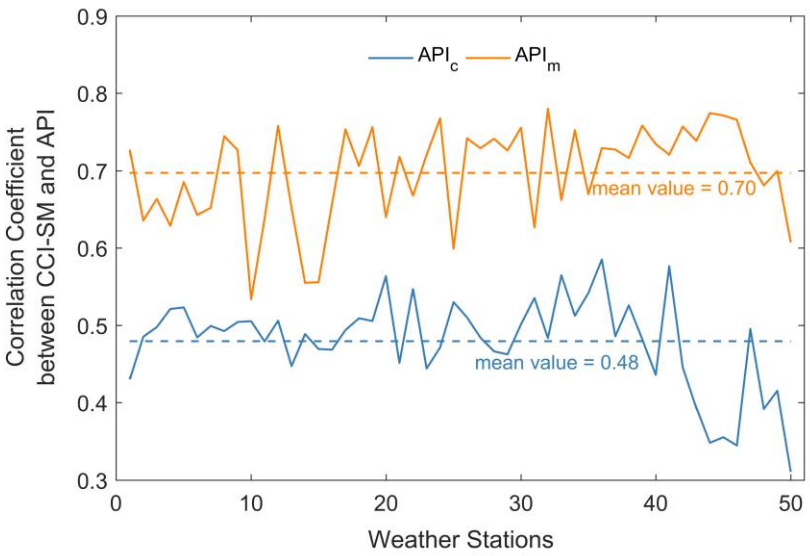

Figure 8. The correlation coefficient between CCI-SM and two versions of API for 50 weather stations as well as their mean values.

4.4 The application of API_m

To further evaluate the performance of API_m in indicating soil moisture conditions, an investigation was carried out to apply it in landslide prediction studies. Here RS threshold was constructed using the rainfall and soil wetness information associated with the occurrence of landslides during the period from 2006 to 2014. For RS threshold, the rainfall indicator is the recent 3-day cumulative rainfall (R) before the landslide occurrence, and the antecedent soil wetness (S) is indexed by the API_m of the day before the recent 3-day. Here the value of API_m is scaled with its minimum and maximum values, ranging from 0 to 1. With the rainfall and soil wetness information of all landslides, RS threshold is determined by various combinations of these two variables' critical values, which are defined at their different percentiles (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 and 50). The percentile distribution of all landslides' antecedent soil wetness and recent rainfall is shown in Figure 9, where the critical values of these two variables are marked with red triangles. From Figure 9a, the antecedent soil condition for more than 80% of landslides is more than 0.68, while more than 50% of landslides have antecedent soil wetness equal to 1. As for the recent 3-day cumulative rainfall in Figure 9b, it is found there are always rainfall before the landslide occurrence, although the rainfall amount varies a lot. More than 50% of landslides have the recent 3-day cumulative rainfall greater than 36 mm. The big difference of the antecedent soil wetness and 3-day cumulative rainfall of landslides is the reason why various percentiles are investigated.

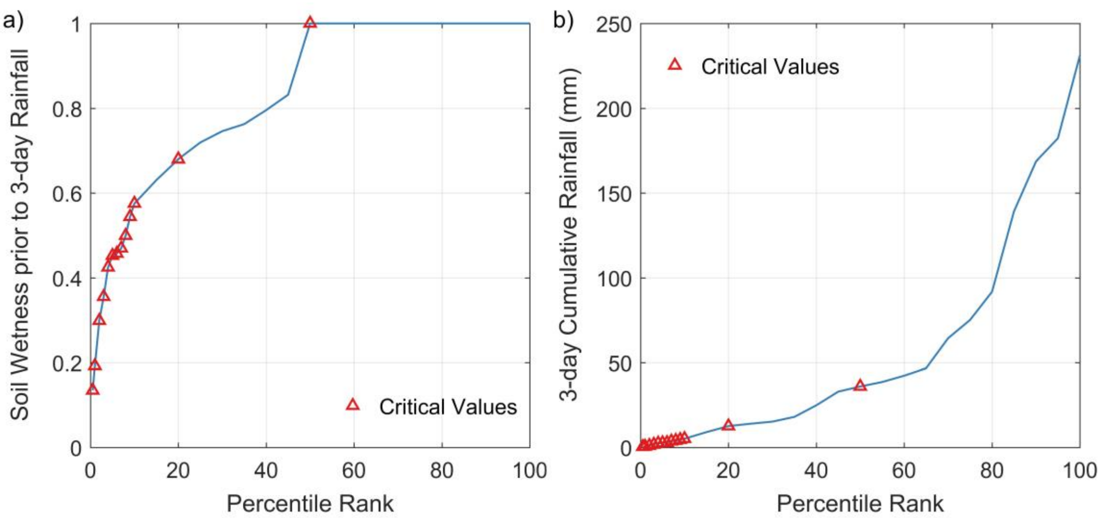

Figure 9. The percentile distribution of all landslides' antecedent soil wetness and recent 3-day cumulative rainfall (soil wetness is indexed with the scaled value of API_m).

The prediction performance of different RS thresholds was evaluated using the data from 2015 to 2016, whose ROC curves are shown in Figure 10. In this figure, the point represents one scenario of RS threshold, determined by one combination of the critical value of soil wetness and rainfall. The critical value of soil wetness remains the same for the points on the same curve, for instance, the critical value of soil wetness is determined with its 1st percentile (P1) for the points on the red curve. The difference among the points on the same curve is the rainfall's critical value, from right to left, the rainfall's critical value is determined by 12 different percentiles at the percentile rank of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 and 50. It can be seen from Figure 10, when the critical value of the antecedent soil wetness remains the same, increasing the critical value of the recent 3-day cumulative rainfall could improve the false alarm rate sometimes at the expense of reducing the hit rate. This result also applies to the case in which the rainfall's critical value remains the same and the soil wetness' critical value increases. In order to determine the optimal critical value of the antecedent soil wetness, the area under the ROC curve (AUC) is employed, the larger the area, the better the prediction performance. Based on the value of AUC, the critical value of the soil wetness determined with the 10th percentile could provide the best predictive capabilities (relative higher hit rates and lower false alarm rates), which is used to compare with the rainfall threshold's performance.

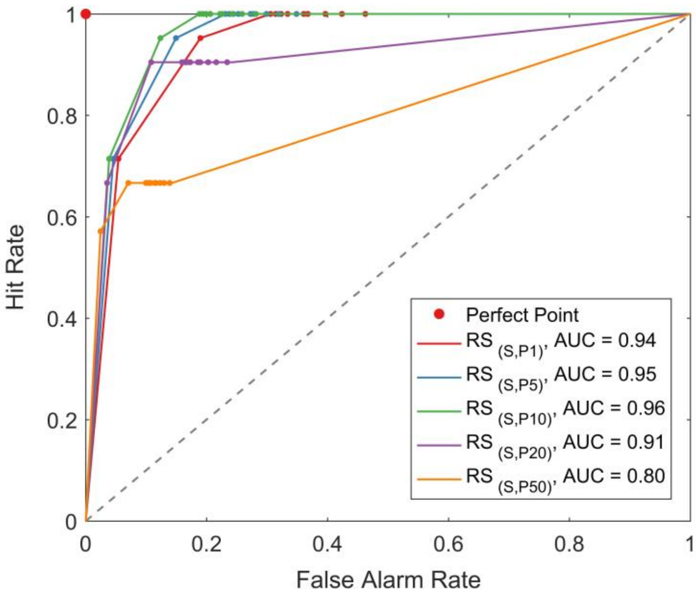

Figure 10. Receiver operator characteristic (ROC) curves for various RS thresholds with the area under the curve (AUC) listed (" S, Pi " means the critical value of the antecedent soil wetness is determined with the ith percentile).

The rainfall thresholds with different percentile ranks were determined for comparison (Table 1a). Three thresholds with the 1st, 5th and 50th percentiles are shown in Figure 11 as well as the rainfall conditions (D, E) that are likely to trigger landslides. Rainfall conditions associated with landslides are in the range of duration $1\ day \leq D \leq 50\ days$, and in the range of cumulative event rainfall $9.6mm \leq E \leq 637.2mm$, which are the ranges of validity for the threshold. Taking the percentile rank of 5 as an example, as expected, there are 5 pairs of the (D, E) data (5% of 112 rainfall conditions) below the P5 threshold. It is noted that the uncertainties of the thresholds depend on the number and distribution of the empirical data, and increasing the sample size could reduce the uncertainties.

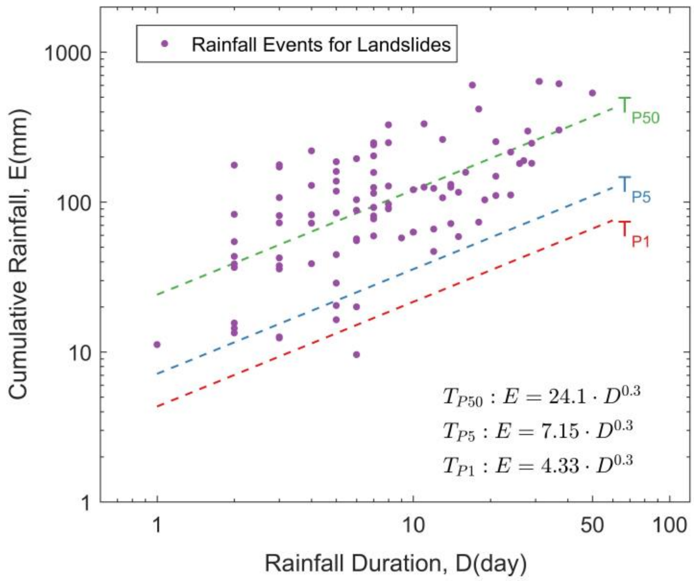

Figure 11. The rainfall thresholds with percentile ranks of 1, 5 and 50, as well as the rainfall conditions (D, E) that are likely to trigger landslides.

The ROC curves are plotted in Figure 12 for rainfall thresholds and RS thresholds whose soil wetness's critical value is determined with the 10th percentile. The statistical indicators of those thresholds are summarized in Table 1. From Figure 12, it is clear that the false alarm rate is greatly reduced by the RS threshold, leading to a higher value of AUC than the rainfall threshold. To determine the optimal threshold that meets a balance between the hit rate and the false alarm rate, Euclidean distance of various thresholds is compared. The rainfall threshold with the smallest distance to the perfect point is achieved at the percentile rank of 20, where HR is 0.818 and FAR is 0.353. RS threshold has the smallest distance when the soil wetness's critical value is defined with its 10th percentile and the rainfall's critical value is defined with its 20th percentile, whose HR is 0.955 and FAR is 0.124. Furthermore, the optimal threshold is also determined by restricting HR as 1 due to the danger of missed alarms. In this way, the rainfall threshold has the best performance for the 3rd percentile case with FAR as 0.698, while the optimal RS threshold is achieved at the 10th percentile for both soil wetness and rainfall component, with FAR is 0.188. Through comparing these two types of the optimal thresholds, it is found that the prediction performance of the optimal RS thresholds is closer to the perfect point, compared with the optimal rainfall thresholds, indicating the application of API_m in landslide prediction studies is effective by indicating antecedent soil moisture conditions of the landslide occurrence.

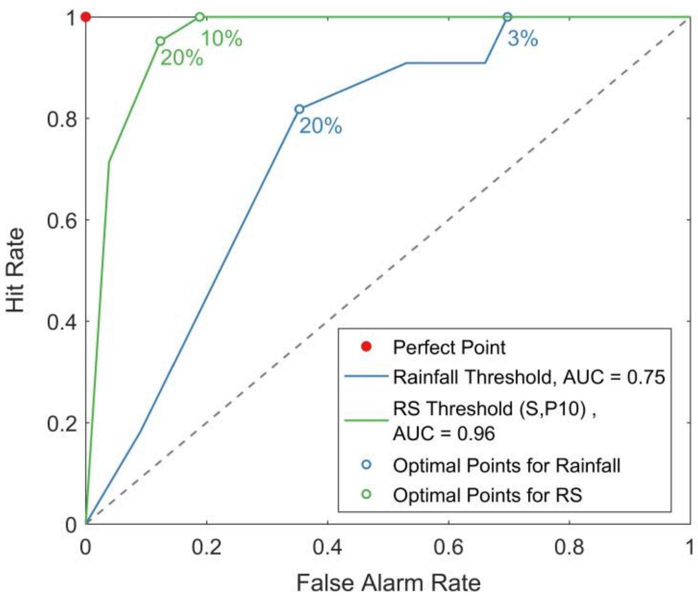

Figure 12. Receiver operator characteristic (ROC) curves for rainfall thresholds and RS thresholds whose soil wetness's critical value is determined with the 10th percentile, as well as the area under the curve (AUC) listed.

Table 1. The prediction results for various thresholds in terms of TP, FN, FP, FN, HR, FAR and the Euclidean distance (d). The optimal results are shown in bold.

a) Rainfall thresholds

P	$E = \alpha \cdot D^{\gamma}$		TP	FN	FP	TN	HR	FAR	d
	α	γ							
1	4.33	0.30	22	0	612	172	1.000	0.781	0.781
2	5.29	0.30	22	0	575	209	1.000	0.733	0.733
3	6.01	0.30	22	0	547	237	**1.000**	**0.698**	**0.698**
4	6.61	0.30	20	2	518	266	0.909	0.661	0.667
5	7.15	0.30	20	2	499	285	0.909	0.636	0.643
6	7.64	0.30	20	2	475	309	0.909	0.606	0.613
7	8.11	0.30	20	2	462	322	0.909	0.589	0.596
8	8.54	0.30	20	2	444	340	0.909	0.566	0.574
9	8.96	0.30	20	2	428	356	0.909	0.546	0.553
10	9.35	0.30	20	2	416	368	0.909	0.531	0.538
20	12.94	0.30	18	4	277	507	**0.818**	**0.353**	**0.397**
50	24.10	0.30	4	18	71	713	0.182	0.091	0.823

b) RS thresholds

P for R	Threshold value		TP	FN	FP	TN	HR	FAR	d
	S(P10)	R (mm)							
1	0.58	0.58	22	0	2456	6267	1.000	0.282	0.282
2	0.58	1.03	22	0	2256	6467	1.000	0.259	0.259
3	0.58	1.72	22	0	2114	6609	1.000	0.242	0.242
4	0.58	2.45	22	0	1974	6749	1.000	0.226	0.226
5	0.58	2.60	22	0	1971	6752	1.000	0.226	0.226
6	0.58	2.74	22	0	1937	6786	1.000	0.222	0.222
7	0.58	3.63	22	0	1796	6927	1.000	0.206	0.206
8	0.58	4.01	22	0	1737	6986	1.000	0.199	0.199
9	0.58	4.58	22	0	1691	7032	1.000	0.194	0.194
10	0.58	5.14	22	0	1644	7079	**1.000**	**0.188**	**0.188**
20	0.58	12.60	21	1	1079	7644	**0.955**	**0.124**	**0.132**
50	0.58	36.00	16	6	338	8385	0.727	0.039	0.275

5 Discussion

The modified version of API proves to be more correlated with the observed soil moisture compared with the conventional version, as it is more in line with the physical process. Since the formulation of the modified API is simple and the input data is less demanding than the commonly used hydrological models, it is easier to be used in practical issues. One major challenge with its application is the determination of parameters. In this study, the calibration method is used to estimate the parameters. However, this approach is only applicable to sites with the observed soil moisture data. For sites with the in-situ measured soil moisture with limited temporal coverage, the calculation of API_m with the calibrated parameters would be useful for the data extension, as the temperature and rainfall data are always available for a long-term period. For sites without in-situ measurements of soil moisture, although the parameter extrapolation is an approach to help estimate parameters, the performance of parameters estimated in this way varies greatly, which can be seen from Figure 8. Using the satellite soil moisture as a proxy of the in-site measured soil moisture to calibrate parameters could be another way, however, the uncertainties associated with the remote sensed data may affect the parameter calibration. For the purpose of better applying API_m to practical issues, the determination of the parameters needs further exploration. For instance, when the in-situ measurements of soil moisture are sufficient, some relationships could be constructed between the calibrated parameters and characteristics of different locations. In this way, the parameters could be extrapolated to ungauged sites based on these relationships, thus the application of API_m will not be restricted by the determination of parameters.

The API_m provides a useful and effective indicator for soil moisture conditions. Although it is not able to estimate the absolute soil moisture value, its good correlation with the observed soil moisture data demonstrates it could capture the temporal evolution of soil moisture conditions. As a result, using API_m as an indicator of soil moisture conditions could provide a proxy of soil moisture for some applications, where only the general soil moisture conditions are required rather than the absolute soil water content values. The landslide early warning is one of those applications, in which API_m could be used as an index of the antecedent soil moisture conditions before landslide occurrence. Moreover, for those explorations that investigate the effect of the soil moisture conditions on the catchment runoff response, API_m is also useful.

In this study, an investigation on the application of API_m in landslide studies was performed. The API_m dataset for the landslide-prone area was derived with the parameters calibrated in the San Pietro Capofiume site, although the parameter extrapolation shows different reliability for different locations, the API_m's correlation with the CCI-SM dataset is generally acceptable, since there are uncertainties associated with the satellite soil moisture data. In this application, API_m was used to index antecedent soil moisture conditions, and then employed to construct RS threshold which consists of the recent 3-day rainfall component and the antecedent soil wetness component. Direct comparison between RS threshold and rainfall threshold confirms that RS threshold is able to provide better prediction capabilities in terms of higher hit rate and lower false alarm rate. One possible reason for this improvement is that RS threshold takes into account the antecedent soil moisture conditions, which are widely recognized as the important factor in the initiation of landslides. As a result, it is inferred that API_m is feasible to indicate the soil moisture condition. It is noted although RS threshold considers antecedent soil moisture conditions, the only required data is the rainfall data, therefore, RS threshold used in this study could facilitate the integration of the rainfall forecasts and fulfil predicting landslide occurrence in advance for the following day.

6 Conclusion

In this study, API is modified by incorporating more considerations of the hydrological process. First, the recession coefficient is allowed to vary with the change of temperature, and thus more consistent with the physical process of water loss. Second, the value of API is restricted by a maximum value of API to avoid overestimating the soil moisture. The added parameters are determined and validated by comparing the API_m dataset with the in-situ measured soil moisture. The better correlation between these two datasets demonstrates that API_m could better indicate the soil moisture condition, compared with API_c. This capability was further explored by applying API_m to landslide prediction studies. The API_m was calculated for the landslide-prone area in the Emilia-Romagna region, northern Italy, which is used to indicate the antecedent soil moisture condition when constructing RS thresholds. RS threshold shows an improved prediction performance than the rainfall threshold, with a higher hit rate and lower false alarm rate. This improvement was the result of accounting for the antecedent soil moisture condition, further indicating the validation of API_m to indicate the soil moisture condition.

The results reported here demonstrate the effectiveness of the modifications proposed to the conventional version of API, which improves the performance of API to indicate the soil moisture condition. Although the parameters of the modified API need to be calibrated for different locations, the simple formulation and easy availability of the required data make it more practical in some applications, through providing an effective proxy of soil moisture. In order to better apply the modified API to practical issues, more explorations are encouraged to test and improve its performance.

Acknowledgements

The authors acknowledge Dr. Matteo Berti for providing landslides data and Arpae Emilia-Romagna organization for providing the meteorological data. The first author would like to thank the China Scholarship Council for funding her study in the University of Bristol. This study is supported by the National Natural Science Foundation of China (41871299), Resilient Economy and Society by Integrated SysTems modelling (RESIST) (Newton Fund via Natural Environment Research Council (NERC) and Economic and Social Research Council (ESRC) (NE/N012143/1)) and the Fundamental Research Funds for the Central Universities of China (2016B42014).

References

Bolten JD, Crow WT, Zhan X, Jackson TJ and Reynolds CA (2010) Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 3: 57-66. doi: 10.1109/jstars.2009.2037163

Brocca L, Melone F, Moramarco T, Wagner W, Naeimi V, Bartalis Z and Hasenauer S (2010) Improving runoff prediction through the assimilation of the ascat soil moisture product. Hydrology and Earth System Sciences 14: 1881-1893. doi: 10.5194/hess-14-1881-2010

Brocca L, Morbidelli R, Melone F and Moramarco T (2007) Soil moisture spatial variability in experimental areas of central italy. Journal of Hydrology 333: 356-373. doi: 10.1016/j.jhydrol.2006.09.004

Brunetti M, Peruccacci S, Rossi M, Luciani S, Valigi D and Guzzetti F (2010) Rainfall thresholds for the possible occurrence of landslides in italy. Natural Hazards and Earth System Sciences 10: 447-458.

Castillo V, Gomezplaza A and Martinezmena M (2003) The role of antecedent soil water content in the runoff response of semiarid catchments: A simulation approach. Journal of Hydrology 284: 114-130. doi: 10.1016/s0022-1694(03)00264-6

Chleborad AF, Baum RL, Godt JW and Powers PS (2008) A prototype system for forecasting landslides in the seattle, washington, area. Reviews in Engineering Geology 20: 103-120. doi: 10.1130/2008.4020(06)

Crow W, Bindlish R and Jackson T (2005) The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff partitioning. Geophysical Research Letters 32.

Crozier M and Eyles R (1980) Assessing the probability of rapid mass movement. Third Australia-New Zealand conference on Geomechanics: Wellington, May 12-16, 1980, Institution of Professional Engineers New Zealand, pp 2

Draper CS, Walker JP, Steinle PJ, de Jeu RAM and Holmes TRH (2009) An evaluation of amsr–e derived soil moisture over australia. Remote Sensing of Environment 113: 703-710. doi: 10.1016/j.rse.2008.11.011

Entekhabi D, Reichle RH, Koster RD and Crow WT (2010) Performance metrics for soil moisture retrievals and application requirements. Journal of Hydrometeorology 11: 832-840. doi: 10.1175/2010jhm1223.1

Fedora MA (1987) Simulation of storm runoff in the oregon coast range.

Gariano SL, Brunetti MT, Iovine G, Melillo M, Peruccacci S, Terranova O, Vennari C and Guzzetti F (2015) Calibration and validation of rainfall thresholds for shallow landslide forecasting in sicily, southern italy. Geomorphology 228: 653-665. doi: 10.1016/j.geomorph.2014.10.019

Glade T, Crozier M and Smith P (2000) Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "antecedent daily rainfall model". Pure and Applied Geophysics 157: 1059-1079.

Gruhier C, Rosnay Pd, Hasenauer S, Holmes T, De Jeu R, Kerr Y, Mougin E, Njoku E, Timouk F and Wagner W (2009) Soil moisture active and passive microwave products: Intercomparison and evaluation over a sahelian site.

Ibsen M-L and Casagli N (2004) Rainfall patterns and related landslide incidence in the porretta-vergato region, italy. Landslides 1: 143-150.

Jackson TJ, Cosh MH, Bindlish R, Starks PJ, Bosch DD, Seyfried M, Goodrich DC, Moran MS and Du J (2010) Validation of advanced microwave scanning radiometer soil moisture products. IEEE Transactions on Geoscience and Remote Sensing 48: 4256-4272. doi: 10.1109/tgrs.2010.2051035

Kerr YH, Waldteufel P, Wigneron J-P, Delwart S, Cabot F, Boutin J, Escorihuela M-J, Font J, Reul N, Gruhier C, Juglea SE, Drinkwater MR, Hahne A, Martín-Neira M and Mecklenburg S (2010) The smos mission: New tool for monitoring key elements ofthe global water cycle. Proceedings of the IEEE 98: 666-687. doi: 10.1109/jproc.2010.2043032

Koster RD, Dirmeyer PA, Guo Z, Bonan G, Chan E, Cox P, Gordon C, Kanae S, Kowalczyk E and Lawrence D (2004) Regions of strong coupling between soil moisture and precipitation. Science 305: 1138-1140.

Koster RD, Guo Z, Yang R, Dirmeyer PA, Mitchell K and Puma MJ (2009) On the nature of soil moisture in land surface models. Journal of Climate 22: 4322-4335. doi: 10.1175/2009jcli2832.1

Koster RD, Mahanama SPP, Livneh B, Lettenmaier DP and Reichle RH (2010) Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. Nature Geoscience 3: 613-616. doi: 10.1038/ngeo944

Linsley RK, Kohler MA and Paulhus JL (1949) Applied hydrology. The McGraw-Hill Book Company, Inc.; New York,

Martelloni G, Segoni S, Fanti R and Catani F (2011) Rainfall thresholds for the forecasting of landslide occurrence at regional scale. Landslides 9: 485-495. doi: 10.1007/s10346-011-0308-2

Melillo M, Brunetti MT, Peruccacci S, Gariano SL and Guzzetti F (2014) An algorithm for the objective reconstruction of rainfall events responsible for landslides. Landslides 12: 311-320. doi: 10.1007/s10346-014-0471-3

Mirus BB, Becker RE, Baum RL and Smith JB (2018) Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning. Landslides 15: 1909-1919. doi: 10.1007/s10346-018-0995-z

Posner AJ and Georgakakos KP (2015) Soil moisture and precipitation thresholds for real-time landslide prediction in el salvador. Landslides 12: 1179-1196. doi: 10.1007/s10346-015-0618-x

Rossi M, Witt A, Guzzetti F, Malamud BD and Peruccacci S (2010) Analysis of historical landslide time series in the emilia-romagna region, northern italy. Earth Surface Processes and Landforms 35: 1123-1137. doi: 10.1002/esp.1858

Scheevel CR, Baum RL, Mirus BB and Smith JB (2017) Precipitation thresholds for landslide occurrence near seattle, mukilteo, and everett, washington. Open-File Report, Reston, VA, pp 60

Staley DM, Kean JW, Cannon SH, Schmidt KM and Laber JL (2013) Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern california. Landslides 10: 547-562.

Valenzuela P, Domínguez-Cuesta MJ, Mora García MA and Jiménez-Sánchez M (2017) Rainfall thresholds for the triggering of landslides considering previous soil moisture conditions (asturias, nw spain). Landslides 15: 273-282. doi: 10.1007/s10346-017-0878-8

Wagner W, Naeimi V, Scipal K, de Jeu R and Martínez-Fernández J (2006) Soil moisture from operational meteorological satellites. Hydrogeology Journal 15: 121-131. doi: 10.1007/s10040-006-0104-6

Zêzere J, Vaz T, Pereira S, Oliveira S, Marques R and Garcia R (2015) Rainfall thresholds for landslide activity in portugal: A state of the art. Environmental Earth Sciences 73: 2917-2936.

Zêzere JL, Trigo RM and Trigo IF (2005) Shallow and deep landslides induced by rainfall in the lisbon region (portugal): Assessment of relationships with the north atlantic oscillation. Natural Hazards and Earth System Science 5: 331-344.